DOCUMENT RESUME

ED 399 961 IR 056 104

AUTHOR	Jian, Ming
TITLE	Cataloging Internet Resources.
PUB DATE	96
NOTE	21p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Access to Information; Bibliographic Records; *Cataloging; Higher Education; Information Retrieval; *Internet; Librarians; *Library Catalogs; *Online Catalogs; Online Searching; Technological Advancement; World Wide Web
IDENTIFIERS	Electronic Media; OCLC; University of Nebraska Lincoln

ABSTRACT

Internet cataloging facilitates access and adds value to Internet resources by helping the user find needed resources with relative ease and in a timely manner. It attests to the value of traditional library principles of organizing information and the value of catalog librarians' experience and skills. This paper explores the suitability and applicability of current cataloging rules and tools for Internet resources, and indicates how traditional cataloging methods should evolve. Creating, organizing, and providing access points for Internet resources, and maintaining the currency, persistence, and validity of an Internet catalog present serious challenges for catalog librarians. The paper also discusses OCLC's InterCat Project, an online catalog that uses traditional cataloging standards, the University of Nebraska-Lincoln's (UNL) Internet Resources Catalog, other library World Wide Web databases, providing Web access through the online public access catalog, and Internet catalog maintenance. Six figures show an OCLC InterCat Record, UNL HTML Work-Form; UNL Internet Catalog Record; UNL Internet Catalog Database; UNL Internet Catalog Search Mode; and OCLC's PURL (Persistent Uniform Resource Locator). (Contains 21 references.) (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 399 961

CATALOGING INTERNET RESOURCES

Ming Jian, UNL

Abstract

Internet cataloging facilitates access and adds value to Internet resources by helping the user find needed resources with relative ease and in a timely manner. It attests to the value of traditional library principles of organizing information and the value of catalog librarians' experience and skills. Internet Cataloging explores the suitability and applicability of current cataloging rules and tools for Internet resources, and indicates how traditional cataloging methods should evolve. How to create, organize, and provide access points for Internet resources, and how to maintain the currency, persistence and validity of an Internet catalog present serious challenges for catalog librarians.

I. Should Internet Resources Be Cataloged?

In recent years, we have witnessed a significant and increasing amount of valuable information resources available through the Internet. Electronic scholarly publishing is flourishing, and the explosive growth of World Wide Web resources is creating a vast new source of potential knowledge. It has become clear to the library community that librarians need to play an active role in facilitating access to Internet resources and in organizing Web information. Indeed, a growing number of libraries are organizing Internet resources and making them available to their patrons in various ways, including Internet catalogs.

Librarians started to catalog Internet resources a few years ago. But it is OCLC's InterCat project beginning in 1994 that significantly fostered the awareness of catalog librarians' efforts in this scenario. Internet cataloging has generated both enthusiasm and skepticism. The main arguments against Internet cataloging include:

1) A good number of powerful Internet search tools and comprehensive

IR056104

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ming Jian
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

automatic indexes already exist. Therefore, Internet cataloging is superfluous;

2) Internet resources are ephemeral, unstable, and qualitatively uneven, thus their academic value is questionable;

3) Cataloging is labor-intensive and requires professional expertise. In light of their ephemeral nature and uneven quality, Internet resources are not worth the cataloging efforts.

It is true that many powerful Internet search tools are already available, and their number is growing rapidly. We are all familiar with the Gopher search tool Veronica, and Web search engines and indexing tools such as Lycos, WebCrawler, World-Wide Web Worm, Alta Vista, etc. These search tools feature a wide variety of search interfaces and capabilities, and many of them are comprehensive, dynamic, and accessible all the time at no charge.

Nevertheless, the biggest problem of these search engines is that the comprehensive and dynamic resource pool is not well organized. As a result, they usually generate too many false hits and too much noise in response. Search results are often redundant, unpredictable, sometimes misleading, and for the most part not annotated. Consequently, despite the rapid proliferation of Internet search tools, Internet users are still overwhelmed with the information received and find it very difficult to locate needed information resources. A survey of Internet use through the University of Toronto Library shows that 46% of users did not find anything sought, only 9% found exactly what they were looking for (Tillotson, 1995). Information overflow and the difficulty of finding needed information have become a serious problem in the information age. This is a problem which cannot be solved by Web robots and automatic indexes; and it is an area where librarians can make their contributions with their skills, expertise, and the principles developed by them to organize Information during the last century-and-a-half.

Most Internet search tools rely on automated means to identify, index, and organize Internet resources. Unfortunately, automated means are not able to produce satisfying

search results for Internet users and will not be able to in the near future. Many useful access points "must be provided by people -- people who have a thorough grounding in the principles of information organization" (Taylor, p.631). The main reason automatic means cannot successfully organize information is "that for the most part, language and ideas are simply too ambiguous for automated retrieval systems to properly identify and evaluate. It appears that artificial intelligence technologies will not meet this challenge in the near future, so intellectual labor is necessary to provide qualitative assessment of the Internet's information" (Clearinghouse).

Cataloging is a process of using intellectual labor to analyze and organize the materials cataloged. Unlike automatic indexes, which function as the very basis of information retrieval and usually only provide a simple list of title and keywords used in the documents, cataloging describes the item, analyzes its content, classifies it, assigns subject headings for it, and creates appropriate access points including the necessary authority control.

Compared to automatic indexes and Internet search engines, Internet cataloging has several advantages: 1) with its subject analysis, Internet cataloging strongly supports a browsable, hierarchical subject arrangement of the Internet database. This is very helpful for information seekers in an academic environment, where subject searching is the predominant mode of searching. Research suggests that subject search accounts for more than one-half of all searches (Mischo, p.47). Also, with the subject search, the user will often benefit from the serendipitous discovery of related items; 2) authority control brings together authors and titles, thus adding access points to the item cataloged; 3) it provides more information about the item, including annotations and content notes, which enable the user to determine if the item is useful, and to choose the few items he wants from amongst millions. In summary: Internet cataloging reduces false hits and produces a higher relevancy rate of items retrieved, provides extra access points to the items, enables Internet users to find needed resources with relative ease and in a timely manner, and thus facilitates access and adds value to Internet resources.

It is also true that Internet resources are unstable: they do not have permanent network presence, and URLs (Uniform Resource Locator) are always subject to change. Some Internet resources are not properly preserved or archived; they may disappear after existing on the Internet for a while, which makes their academic value questionable. As we do not own the information sources cataloged in many cases, we do not have control and cannot count on them. In addition to their ephemeral nature, quantitatively immense Internet resources are qualitatively uneven; some are good, some are bad, some turn out to be really ugly. Giving all these factors, it is legitimate to ask whether it is necessary and worthwhile to catalog Internet resources, which is an expensive and time-consuming effort requiring professional expertise.

We believe that Cataloging Internet resources is necessary and worthwhile, because existing Internet search tools and automatic indexes fail to provide Internet users with a relatively easy and timely way to find needed resources on the Internet; and as discussed above, Internet cataloging facilitates access and adds value to Internet resources. Also, Internet resources constitute a big proportion of information sources available today, and some of them do have high research and educational values. Library community has a role to play in organizing Internet resources, and Internet cataloging is one of the important contributions we can make. Hence, the question is not whether Internet resources should be cataloged or not, but how to catalog them; how to choose useful and valuable materials to catalog; how to keep catalog records accurate, persistent and reliable; how to work with information and computing communities to assure that Internet resources have permanent network presence at fixed or virtual locations; how to best preserve Internet resources; and how to make cataloging Internet resources less labor intensive and economically more affordable. These are serious challenges for catalog librarians and the whole library community. Whether Internet cataloging is worthwhile very much depends on how we meet these challenges.

II. Selecting Internet Resources to Catalog

Unlike most automatic indexing, which takes whatever is submitted or found by Web robots, cataloging Internet resources starts with carefully identifying and selecting materials to catalog. Librarians create value by selecting what is valuable and appropriate to their users. At UNL, selecting Internet resources has become a part of library's collection development. Internet resources are selected by librarians, and primarily by the subject specialists who are in charge of collection development. Their experience and subject expertise help them successfully evaluate and select Internet resources. The coherent rationale for selecting Internet resources is based on faculty and student information needs. Conventional collection development criteria are applied consistently across formats including Internet resources.

Just as with print materials, we must be selective. There are an overwhelming number of Internet resources. They constitute the biggest chunk of information sources we have ever tried to tackle. A sizable amount of these information sources have very little or no value at all. Being selective and critical not only ensures the quality of our collection, but is also important to keep cataloging Internet resources economically more affordable, because creating quality cataloging records is expensive.

Besides the consideration of intellectual content, we prefer cataloging well established sites and local networked resources. Well established sites have more stability and are less likely to disappear from the Internet. Local networked resources such as the networked resources created by the university community and the local government information resources are almost always valuable for our clientele. It is also easier for us to identify and communicate with the resource provider. This will help us better keep track of the changes of information sources.

III. Different Cataloging Approaches

During the last century-and-a-half, catalog librarians have developed principles, rules, and formats to organize print materials, and materials on microform, video, etc. Their cataloging principles and tools have proven to be very successful in bringing order out

of the chaos of information. Typically, they provide name, title, subject and other useful access points to information resources cataloged. They also provide cross references for subjects and authority linkage. All these traditional library approaches are still useful for cataloging Internet resources.

However, cataloging methods should be determined ultimately by patrons' information needs, the object being cataloged, and the typical model of information access. Changes in these fields will call for changes in cataloging methods and tools. In recent years, the emergence and the explosive growth of electronic information, especially the Internet, have dramatically changed the way electronic and Internet information resources are sought, accessed and delivered. The remote accessibility of Internet resources, for example, has "profound consequences for bibliographic access and catalog design" (Buckland, p.48). Since Internet resources are remotely accessible, and since libraries do not own the information sources cataloged in most cases, the traditional function of a catalog as a guide to local holdings disappears. Now that the catalog of Internet resources not only provides bibliographic access to the information sources, but also the information sources themselves, new information usually found in a bibliography such as abstracts and annotations has become very useful for Internet catalog.

As of yet there has been no widely accepted standard format, standard data structure, and standard method and procedure for Internet cataloging. Each library and institution takes a different approach to cataloging Internet resources. On the positive side, this situation allows librarians to creatively explore the best way to use their thorough grounding in the principles of information organization, their knowledge, skills, and experience to organize Internet resources.

OCLC's InterCat Project

In October 1994, OCLC started a U.S. Department of Education funded 18-month project called "Building a Catalog of Internet Resources". As of June 14, 1996, about

206 libraries and institutions have participated in the project and contributed 4494 MARC records, which are now searchable via the OCLC Online Union Catalog and via the InterCat database (http://www.oclc.org:6990). This is probably the most traditional library approach to providing access to Internet resources, because the catalog records (Figure I) fully comply with traditional cataloging formats and rules: USMARC format, AACR2, and LC subject headings. MARC field 856 "Electronic location and access" is used to "locate an electronic item', and to "identify the electronic location containing the items or from which resource is available" (Olson, 1995).

Using AACR2 and MARC format for Internet cataloging has its advantages. AACR2 has been the national cataloging standard for a long time, and has proven to be successful as the standard for cataloging print materials and materials in other formats. As a machine-readable cataloging standard, MARC, along with the standard communication provided by the International Standards Organization (ISO), has been a common shareable basis of bibliographic data and data structure in an online environment. For the library community, AACR2, MARC, and LCSH are the backbone of bibliographic control, the primary tools for precise information retrieval, and common standards for the extensive sharing of data and cooperative cataloging. Using these widely accepted pre-existing standards will not only help the standardization of bibliographic data and data structure, but also help integrate Internet resources with other long-existing library materials and other information resources.

However, using AACR2 and MARC format for cataloging Internet resources also has some significant drawbacks. Prepared primarily for printed materials in the 1980s when Internet resources were barely known for most librarians, AACR2 does not provide necessary rules for cataloging networked resources, especially Web resources. We don't blame AACR2 for not providing rules suitable for cataloging Internet resources, because there was no need for that at the time the rules were set up. But the rules must evolve in response to the changes of cataloging objects.

Unfortunately, the library community has not taken action soon enough. For example, the concept of main entry and added entry has long become unimportant or even useless in an online environment (For more on recent discussions, see Bierbaum, 1994). But catalogers still waste time and energy to determine main entry and added entries. Also, it seems that there is no need to devote so much time and attention to determine some details such as the order of entering note fields for Internet resources, which may not make any difference to the patron. Library catalogs used to be centered on the physical objects contained in a library. As a result, the structure of AACR2 clearly emphasizes the description. For most electronic texts, especially Internet resources, however, many physical characteristics of a file are no longer essential.

The existing rules in Chapter 9, "Computer Files", of AACR2 are also not especially suitable for electronic texts, as pointed out by Hockey a few years ago (Hockey, 1993). To make things worse, some local online catalog systems do not even support some new fields and subfields in a MARC record, which we create especially for Internet resources.

As an ANSI standard for information interchange, MARC has been the best tool for bibliographical control. With its high level of standardization, MARC format is a major accomplishment of the bibliographic community. MARC, however, does not satisfy many of the needs for Internet cataloging. Polowitch and Horowitz believe "the most significant reason for this is that it cannot be flexibly applied to information other than that for which it was designed (which, technically, was to produce printed cards for library catalogs)" (Palowitch, p.112). There is no doubt that we need to change or enhance MARC format to make it more flexible and more suitable for Internet resources. It is also necessary and beneficial to bring MARC and other standard formats and data structures such as SGML and EDIFACT together.

According to the manager of the OCLC InterCat project, the main reason OCLC chose MARC format and AACR2 for Internet cataloging is not because of "the

relative merit of MARC and AACR2", but because the project wants to explore "their applicability to remotely accessible electronic resources". An alternative approach would take much more time and thus "seems less reasonable ... than undertaking a limited, short-term, intensive project such as the OCLC Internet Cataloging project" (Jul, 1995).

The library community holds different views on OCLC's InterCat Project (One of the sharp critic is Pat Ensor, cf. Ensor, 1995). I do think that we need national utilities such as OCLC to coordinate our efforts in organizing Internet resources. In my opinion, the project has done at least two things absolutely right: 1) the project is based on the belief that "libraries have a significant contribution to make" to facilitate access to Internet resources (Jul, 1996); and 2) the project provides a test bed for catalog librarians to explore, investigate, and test theories and practices of cataloging Internet resources. And the InterCat listserver (Intercat@oclc.org) has become an important forum to discuss issues concerning Internet cataloging .

UNL's Internet Resources Catalog

UNL is among the first libraries to make efforts to organize Internet resources and develop a catalog database of Internet resources. We started to catalog Internet resources and build a Gopher-based Internet Resources Catalog in 1993 (For more on the Gopher cataloging and the early development of UNL libraries' Internet research Gopher, see Swann 1995). This year, we have completed the migration of our Internet Catalog database from Gopher to Web environment. More than 700 Internet resources with educational and research value have been cataloged and made available through our WWW database (http://libfind.unl.edu:2020/home.html).

Our Internet cataloging follows national cataloging standards such as AACR2 and LC subject headings in a flexible manner; in some fields we go beyond them, and in others we do less. Our emphasis is put on providing more useful access points and more information that could help the patron find needed resources with relative ease

and in a timely manner. For example, we go beyond AACR2 in adding content notes and annotations to the catalog records. Also, to enhance subject access, we use not only controlled vocabulary such as LC subject headings and cross-references for subjects in the keywords field, but also natural language in the content notes field to provide annotations and summary notes. All these fields are searchable in our WWW catalog database. This kind of information helps the patron to determine whether the resource is relevant, and retrieve the needed information from among millions of records.

We intentionally avoid unnecessary or unimportant information and some detailed bibliographical descriptions, which are time consuming to create and do not add value to information access. Given the ephemeral nature of Internet addresses, we do not think it is worth our time and energy to provide the same level of description for Internet resources that we expect for print materials.

Unlike OCLC's InterCat project, we do not use MARC format for our Internet cataloging. Instead we created our own inhouse-workform. At first, we created a Gopher workform for cataloging Gopher materials. When the WWW became the predominant Internet tool, we then created a HTML (Hypertext Markup Language) workform (Figure II) for cataloging different kinds of Internet resources. These workforms allow us to catalog Internet resources and maintain the Internet catalog database in a time-efficient manner. At the same time, the HTML workform also makes the Internet cataloging easier for those who do not have HTML experience.

Our Internet catalog records (Figure III) can be created using any text editor or a normal word processor. We take advantage of a Windows environment, which allows multi-tasking to occur in multiple windows, to easily and rapidly transfer useful data from different Internet sites and to copy subject headings from our local online catalog into the catalog records. The finished catalog records are then added to the Internet resource catalog database in batch mode. Included in the database are Web, Gopher, FTP and Telnet sites, listservers, newsgroups, electronic books, texts,

conferences, and journals. They cover most major academic disciplines at UNL, and have high research and educational value.

The UNL Internet resource catalog database provides both browse (Figure IV) and search modes (Figure V). Our database will soon provide a new hierarchical subject index. The new subject hierarchy will cover major academic disciplines and about one hundred subcategories. The patron can easily browse the subject index of the database to find needed information, and enjoy serendipitous discovery of related items in the meantime. The search mode provides not only conventional search capabilities on author, title, subject, keywords, but also the search capability on content notes, which provides extra access points to the materials cataloged. Since natural language is used in the content notes, some of the terms used in this field are more likely to be searched by those patrons who are not familiar with controlled vocabularies. Our search mode also features truncation, nested, and Boolean search capabilities. To connect to the resource found, the patron only needs to click on the highlighted URL in the hyperlinked cataloging records.

Although UNL's Internet catalog differs from traditional library catalog in various ways, it has been created using the same traditional library approach, which underlines carefully evaluating, selecting, describing materials, analyzing them by subject, and providing useful access points to them. It is this approach, not MARC format and AACR2, that is typical traditional library approach.

Other Library Web Databases

The same library approach has also been used by some other better Internet catalog databases or virtual libraries created by librarians. Among those considered top Internet search tools and catalogs are library Web databases such as Clearinghouse for Subject-Oriented Internet Resources Guides, INFOMINE, etc. Clearinghouse for Subject-Oriented Internet Resources Guides (http://www.lib.umich.edu/chhome.html) is developed by the University of Michigan Library and the Library School at the

University of Michigan. It presents a series of more than 400 subject guides in various disciplines, which are accessible from a menu listing broad subject areas. The subject guides are well organized and annotated. INFOMINE (http://lib-www.ucr.edu) is developed by the Library of the University of California, Riverside. It has taken a similar approach to organizing Internet resources as we have done at UNL. INFOMINE provides indexing, annotations and links to Internet resources of scholarly use to the academic community (For more on INFOMINE, see Mitchell, 1996).

These WWW databases could be considered as catalogs "in a wider sense, a list of materials prepared for a particular purpose", as AACR2 defines it (Gorman, p.616). The common point of these databases is their well selected, high-quality resources, their hierarchically structured subject guides, their subject analysis and annotations, and their logical in-depth indexes. These probably represent the fields where librarians can make their best contributions to organize Internet resources.

Compared to robot-based cataloging services such as Lycos and WebCrawler, these Web guides and catalogs contain only a very small portion of networked resources available on the Internet, but they provide well evaluated, carefully selected, and professionally analyzed high quality research and educational Internet resources, which will much better serve the needs of faculty and students in an academic environment. Furthermore, they provide more access points and search capabilities for the materials selected and cataloged.

Widely accepted standard format, standard data structure, and standard method and procedure for cataloging Internet resources are desirable. But before we get that far, we should encourage different experiments and practices. In fact, each different approach mentioned above has made its own contribution to the common goal: making valuable and high quality research and educational Internet resources available to library patrons in an economically affordable way, and making them accessible in an easy and time-efficient manner.

IV. Providing WWW Access through the OPAC

Internet cataloging records can be made accessible for the patron through different channels, such as the catalog of Internet resources with Web or Gopher gateways, OPACs with built-in Web interfaces, catalogs that can be reached through Web to Z39.50 gateways, or even traditional OPACs without Web interfaces. In the last case, the patron may use cut and paste functions in a windows environment to access Web resources. We can either keep Internet catalog records separate, or add them into our local online system and integrate them with existing library materials.

Our goal should be integrating access to Internet resources with access to existing conventional materials in the local online system. Making possible one-stop information shopping would be beneficial to our patrons. Unfortunately, most OPAC systems at this stage are not able to maximize the use of Internet catalog. Also, the instability of Internet resources causes special maintenance problems. Consequently, keeping Internet resources in a separate database might be an alternative at this stage. It could even help the patron retrieve Internet resources in an easier and more efficient way if the patron is only looking for Internet resources. We can then design search capabilities and displays that are specifically geared towards finding Internet resources. For example, the hierarchical subject guide with annotation would be very helpful in searching Internet resources. In addition, separate Internet resource database might be easier to maintain.

In order to better handle Internet resources, OPACs need to expand in scope and to provide enhanced access capabilities, enhanced interface and gateway software. Furthermore, as many Internet catalog records are encoded with HTML instead of MARC format, and more and more electronic texts are tagged and distributed with TEI (Text Encoding Initiative) headers, OPAC systems need to evolve and provide new possibilities to incorporate diverse standard formats and data structures. In this regard, it is encouraging that "the Library of Congress is working with others to develop an SGML version of MARC record that will allow roundtrip compatibility

between the two record structures" (McCallum, p.13). For now, though, catalog records or text files encoded with other formats need to be converted into MARC format in order to be integrated with existing library materials in local systems.

When adding Internet resources to the local online system, we need to take the limitation of the library's online system into consideration and to carefully design the OPAC display of Internet catalog records. Different online library systems download, arrange, and display new fields and subfields (especially 856) in Internet catalog records very differently. Some of the subfields are not supported at all. For example, Innopac only supports $z and $u in 856 field. Some useful information in other subfields such as $3 is simply ignored.

Some OPAC displays of Internet resources could be confusing or misleading for the patron. For example, a call number could be misleading in an Internet catalog record. The patron may go to the stacks and try to find the information under the call number. Libraries who have integrated their Internet catalog with the existing library materials have experienced various display problems in their local systems. One system displays the explicit instruction to "check shelf" for an Internet resource. Another system displays "Not Checked Out" under STATUS. There are also displayed messages indicating that a resource was "in transit", as if on its way to binding, etc. Librarians need to work with local systems vendors to solve these problems.

We may also need to consider what information needs to be displayed and how much information should be provided. For example, should we display system requirements to patrons? If yes, what information and how much? The dilemma for us at this stage is, as somebody pointed out in the InterCat listserver: for those in the know, such explanations are noise. For those who do not understand the terms, they need much more help than we could provide inside a bib record. One solution for this problem may be to direct the patrons who need assistance to the reference desk.

V. Internet Catalog Maintenance

The Internet catalog has its unique maintenance problems. Internet resources are remote and unstable. As we do not own the information sources in most cases, we do not have control on them. URLs are always subject to change, the same URL may have different content at a different time, and some cataloged materials may disappear after existing on the WWW for a while (For more on Web page maintenance see Sha, 1996).

Obviously, the instability of URLs has caused serious maintenance problems of Internet catalog records. As OCLC InterCat project manager Eric Jul puts it: "entering URL into bib records creates a catalog maintenance liability" [Jul, Feb. 1996]. The ultimate solution for the URL problem very much depends on the commitment and the efforts of library, information, and computing communities to assure that Internet resources have permanent network presence at fixed or virtual locations. The establishment of standards for reliable addressing of Internet resources has been addressed through the proposed URN (Uniform Resource Name) resolution of the Internet Engineering Task Force (IETF) (For more on IETF's URN proposals, see IETF URI Working Group). It may be useful to compare URN with ISBN/ISSN and the authority control concept. Like ISBN/ISSN, a URN has global uniqueness, location independence, and persistence. Like an authority file, a URN has an authority name assignment which contains different URLs and meta-information for the authority name. Each Internet resource will be assigned an authority name, which may have potentially many resolution possibilities. When the resolution server receives a client request in URN form, it will check the authority name and return only the current URL that matches the request to the client.

A URN is composed of "Uniform Resource Characteristics (URCs), and Uniform Resource Locators (URLs). URNs are used for identification, URCs for including meta-information, and URLs for locating or finding resources" (Sollins, 1994, 1). The standardization of URN protocol is still very much under development. So far, a

number of schemes have been proposed for assigning and resolving URNs and for associating meta-information with URNs (For a critical review of the proposals, see Madsen, 1995). However, no agreement on a feasible one has been reached.

Before a new standardized resolution protocol, which underlines the standardization of WWW resource identifiers, could be established and widely accepted by the Internet community, what we can do now to keep the catalog records current, accurate and valid is to detect and change invalid URLs on a regular basis. Fortunately, many Web page maintenance tools and URL checkers have been created and widely used to check the status of the URLs in catalog records, such as WebWatch, MOMspider, Netscape SmartMarks, Netbuddy, Red Alert, Webxref, etc. But unfortunately, there is no way to have those Web maintenance tools automatically change the URLs and the related fields in a catalog record. These still have to be done manually.

OCLC has recently implemented PURL (Persistent Uniform Resource Locator) (Figure VI) in its InterCat Project to tackle the persistence problem. Functionally, a PURL is a URL which provides an additional level of indirection. "Instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The resolution service associates the PURL with the actual URL and returns that URL to the client, which can then complete the transaction in the normal fashion" (PURL, 1996). In other words, a PURL may be associated with different URLs, but it always returns the current and valid URL to the client. URLs may change, but PURLs never change.

Although PURLs do not eliminate the instability problem of URLs, they are very useful in a cooperative cataloging environment. Those libraries who download Internet catalog records from OCLC or from cataloging libraries who have created and maintain PURLs, will no longer need to share the catalog maintenance liability and update the URL when it changes. Note that PURL servers will not work automatically, someone has to operate the PURL resolvers. PURLs are not the ultimate solution, but only one component of an incremental solution to the larger

problem.

Conclusion

The library community has a role to play in facilitating access to Internet resources and organizing the Web information. Traditional library principles and approaches in organizing information prove to be valuable in cataloging Internet resources. But cataloging methods, rules, and tools need to be changed in response to new cataloging objects, new models of access, and new information needs of the patron. Cataloging Internet resources presents new challenges to catalog librarians, and many problems still need to be solved. The library community must make a long-term commitment to providing and preserve high quality information resources and access for our patrons, and maintaining the quality and validity of the Internet catalogs.

References

Bierbaum, Esther Green. 1994. *A Modest Proposal: No More Main Entry*. American Libraries. 25(1): 81-84.

Buckland, Michael. 1992. *Redesigning Library Services: A Manifesto*. Chicago: American Library Association.

The Clearinghouse for Subject-Oriented Internet Resource Guides. URL: http://www.lib.umich.edu/chhome.html

Ensor, Pat. 1995. *Libraryland Organizes the Web: An Unnatural Process?* Technicalities. 15(11): 9-11.

Gorman, Michael, Winkler, Paul W., eds. 1988. *Anglo-American Cataloging Rules*. 1988 revision. Chicago: American Library Association.

IETF Uniform Resource Identifiers (URI) Working Group. URL: http://www.ics.uci.edu/pub/ietf/uri

Jul, Erik. 1995. *Reply to Peter Graham's email message: Call for Participation* (May 19, 1995). The email message can be found in: InterCat Listserver Archive. URL: http://ftplaw.wuacc.edu/listproc/intercat/archive.html

Jul, Erik. 1995. *More on PURLs* (Feb 16, 1996). The email message can be found in: InterCat Listserver Archive. URL: http://ftplaw.wuacc.edu/listproc/intercat/archive.html

Jul, Erick. 1996. *Why Catalog Internet Resources?* Computers in Libraries. 16(1): pp.8-10.

McCallum, Sally. 1996. *What Makes a Standard?* Cataloging and Classification Quarterly. 21(3/4):5-15.

Madsen, Mark. 1995. *A Critique of Existing URN Proposals*. IETF URI Working Group Internet-Draft. http://www.ics.uci.edu/pub/ietf/uri/draft-ietf-uri-urn-madsen-critique-00.txt

Mitchell, Steve, Mooney, Margaret. 1996. *INFOMINE -- A Model Web-Based Academic Virtual Library*. Information Technology And Libraries, (March 1996): 20-25. The article is available online at http://lib-www.ucr.edu/pubs/italimine.html

Mischo, William H; Cole, Timothy W. 1992. *The Illinois extended OPAC: library information workstation design and development*. In Ra, Marcha, ed. 1992. Advances in Online Public Access Catalogs. Vol. 1. Westport: Meckler.

Olson, Nancy B. ed. 1995. *Cataloging Internet Resources: A Manual and Practical Guide*. OCLC Online Computer Library Center, Inc. URL: http://www.oclc.org/man/9256cat/toc.htm

Palowitch, Casey, Horowitz, Lisa. 1996. *Meta-Information Structures for Networked Information Resources*. Cataloging and Classification Quarterly. 21(3/4): 109-130.

PURL Frequently Asked Questions. 1996. http://www.oclc.org/purl_faq.html#toc1.2

Sha, Vianne, T., and Timothy B. Patrick, Thomas R. Kochtanek. 1996. *The Traditional Library and the National Information Infrastructure*. URL: "http://www.oclc.org/man/colloq/sha.htm

Sollins, Karen, Larry Masinter. 1994. *Functional Requirements for Uniform Resource Names*. IETF Network Working Group Request for Comments 1737. FTP://ds.internic.net/rfc/rfc1737.txt

Swann, Julie, Rosenquist-Buhler, Carla. 1995. *Developing an Internet research Gopher: Innovation and staff involvement*. 1995. The Journal of Academic Librarianship. 21(5): 371-375.

Taylor, Arlene G. 1994. *The Information Universe: Will We Have Chaos or Control?* American Libraries 25(7): 629-631.

Tillotson, Joy, Cherry, Joan, and Clinton, Marshall. 1995. *Internet Use through the University of Toronto Library: Demographics, Destinations, and Users' Reactions*. Information Technology and Libraries. 14(3): 190-198.

Figure I: OCLC InterCat Record

```
OCLC: 32736424     Rec stat: c
Type:  m   Elvl: I     Srce: d    Audn:     Ctrl:       Lang: eng
Blvl:  m   File: d     GPub:                MRec:       Ctry:  nbu
Desc:  a                                    DtSt: m     Dates: 1992,9999

 1   040    LDL |c LDL
 2   043    n-us-nb
 3   049    LDLL
 4   245 00 Nebraska online |h [computer file].
 5   256    Computer data.
 6   260    Lincoln, Neb. : |b Nebraska Library Commission, |c 1992-
 7   538    Mode of access: Internet. Address: http://www.nol.org
 8   538    System requirements: Internet access, WWW browser.
 9   500    Title from title screen.
10   520    Provides electronic access to state, county, local, association and other
            public information; provides Internet gateway to resources made available
            by state agencies, state funded colleges, and universities, and other
            Nebraska World Wide Web services.
11   650  0 Information networks |z Nebraska.
12   650  0 Internet (Computer network) |z Nebraska.
13   710 2  Nebraska Library Commission.
14   856 7  |3 Nebraska online |u http://www.nol.org |2 http
```

Figure II: UNL HTML Work-Form

```
<HTML>
<HEAD> <TITLE> </TITLE> </HEAD>
<BODY><H4>
ACCESS NUMBER: <P>
SUBJECT: <P>
MENU HEADING: <BR>
<A HREF=http:// > </A> <P>
AUTHOR: <BR> <P>
```

TITLE:
 <P>
KEYWORDS:
<BLOCKQUOTE>
</BLOCKQUOTE> <P>
HTTP SERVER:
 <P>
CONTENT NOTES:
 <P>
</H4> </BODY> </HTML>

Figure III: UNL Internet Catalog Record

MENU HEADING: HTML Converter
AUTHOR: Sendall, Mike
Brandwein, Rich
TITLE: HTML Filters
KEYWORDS:
HTML (Hypertext markup languages)
Hypertext systems
HTML converters
HTML filters
Hypertext Markup Language Converters
HTTP SERVER:
www.w3.org
CONTENT NOTES:
Provides information on converter/filter software between various systems and HTML, including: Word Processor filters, Program Language filters, Man Page, FAQ, Mail filters, Miscellaneous filters, Converting to HTML, Converting from HTML, etc.

Figure IV: UNL Internet Catalog Database

University of
Nebraska-Lincoln
Libraries Internet
Resources Catalog

ABOUT the University of Nebraska-Lincoln Libraries WWW Server
SEARCH the Catalog

INTERNET RESOURCES by Subject:

- **General Reference & News You Can Use**
- **Agriculture & Natural Resources**
- **Art & Architecture**
- **Biological Sciences**
- **Business & Economics**
- **Chemistry**
- **Distance Education**
- **Education**
- **Engineering & Computer Science**
- **Health Sciences**
- **Job Openings from the Internet**
- **Journalism & Communication**
- **Language & Literature**
- **Law, Political Science, & Govt.**
- **Library & Information Science**
- **Performing Arts**
- **Philosophy, Religion & Classics**
- **Physical Sciences**
- **Social Sciences**

Internet Tools:

FTP, Wais, & Gopher..... Electronic Phonebooks..... Archie

NU FRONTIER - ***The University of Nebraska-Lincoln Campus Wide Information System***

BEST COPY AVAILABLE

Figure V. UNL Internet Catalog Search Mode

Item Title: [ ]

RETURN

Searching Instructions

This searches the catalog database on all text including author, title, subject, keywords and content notes. Enter the word you want to search in the above box. You can refine your search with the following:

* ..... is the symbol to truncate a word so that the computer will match on anything that starts with the letters before the *. For example, agri* will match of agriculture or agricultural.

or ..... is used to search for the presence of either word in a file, for example, Britain or British, will bring up all files which have either Britain or British

and .... is used when you only want files that have both words in them, for example, US and military, will bring up files that contain both words, but will **ignore** files that have only military or US

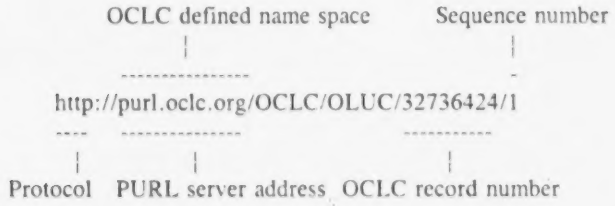
not .... is used when you want to exclude particular words, for example, trees not pine, will bring up files that include the word trees but will **exclude** any files with the word trees that also includes the work pine

" "... is used when you want to find multiple words next to each other, for example "farm workers" will only bring up files that include the two words next to each other, if the file has both words, but they are separated by other text (US farm prices. Workers employed in agriculture) it will not match

( )... parenthesis can be used to nest searches. For example you may want (aids or "immune deficiency syndrome") and (nursing or hospitals)

Figure VI: OCLC's PURL

OCLC's PURL (http://purl.oclc.org)



Display PURL

PURL	http://purl.oclc.org/OCLC/OLUC/32736424/1
URL	telnet://neon.ncl.state.ne.us
URL	http://www.nol.org